Characterizing the expression of individual immune cell types within pediatric central nervous system cancers

Research Thesis

Presented in Partial Fulfillment of the Requirements for Graduation *with Research Distinction* in Molecular Genetics in the College of Arts and Sciences of The Ohio State University

By

Olivia E. Grischow

The Ohio State University

Spring 2021

Project Advisor: Dr. Elaine Mardis, Professor of Pediatrics at The Ohio State University College of Medicine and Co-Executive Director of the Steve and Cindy Rasmussen Institute for Genomic Medicine at Nationwide Children's Hospital

Table of Contents

ABSTRACT

This project defines the utility of bulk RNA sequencing and single-cell RNA (scRNA) sequencing in characterizing the tumor microenvironment of pediatric medulloblastomas. Comparisons between the two methods can identify how they can be used to further define the understudied medulloblastoma. These comparisons were achieved through subtyping based on gene expression data and evaluating the immune expression of primary versus recurrent tumors. Our cohort consists of five medulloblastoma patients that underwent bulk RNA and scRNA sequencing. Bulk RNA sequencing was utilized to accurately subtype medulloblastomas using gene expression data. It was found that scRNA sequencing can further characterize medulloblastomas subtypes through the classification of distinct cell populations between group 3 and group 4 medulloblastomas. Not only is robust subtyping possible, but scRNA sequencing can be used to identify cell types and their differential expression across several samples. Immune expression can also be analyzed through gene expression analysis and scRNA sequencing and is represented in this paper through the longitudinal study of a primary medulloblastoma versus a recurrent medulloblastoma in the same patient. Gene expression analysis using the PanCancer Immune Panel exhibited expression of immune genes as compared to other tumor types and identified under and overexpressed genes. Through scRNA sequencing, it was identified that in one case of a recurrent medulloblastoma, immune cell infiltration increased. These data and futures studies that utilize bulk and scRNA sequencing to characterize tumors can aide in designing immune therapies that give medulloblastoma patients the best possible outcomes.

INTRODUCTION

In recent years, next generation sequencing has greatly increased our knowledge of genomic and transcriptomic alterations in pediatric central nervous system (CNS) tumors likely associated with tumorigenesis (Gröbner et al., 2018) (Ma et al., 2018). Despite these discoveries, there remains a critical need to develop new treatment paradigms that decrease recurrence of brain tumors, as these tumor types remain the leading cause of disease-related death in children (Curtin et al., 2016). Characterizing the tumor immune microenvironment has been at the forefront of many ongoing cancer research and drug development studies, as immunotherapy is considered one of the most promising treatments in cancer medicine today (Sharma & Allison, 2015). Specifically, in pediatric CNS tumors, there is little known about the tumor immune microenvironment, especially given the apparent immune-privileged status of the CNS.

Cancers are comprised of a complex collection of cell types including normal cells, stromal cells, malignant cells, and infiltrating immune cells, and it is this complexity that predisposes patients to inferior prognoses and possible resistance to treatment (Dagogo-Jack & Shaw, 2018). We typically characterize tumors and their infiltrating immune cells from traditional, "bulk" RNA sequencing data, that results from RNA extraction of the tumor mass and therefore includes all cell types. In particular, the gene expression profiles from bulk RNAseq can be used to characterize the different cellular components and in some cases, to provide additional subtyping information that may be pertinent to treatment or prognosis. Immune cell components can be evaluated by computational deconvolution approaches that essentially pattern match gene expression data to the known gene expression profiles for different immune cell types. However, this approach is somewhat limited in that it lacks resolution needed to distinguish all immune cell types and cannot provide information about

important immune cell states, such as T-cell exhaustion or activation. This leads to the omission of critical differences about the immune microenvironment (Venteicher et al., 2017) that may be highly informative to physicians, in terms of likely response to certain therapies. In order to circumvent these challenges, single cell RNA sequencing (scRNAseq) has emerged as a powerful technique that reveals gene expression profiles of individual cells from dissociated tumor tissue and can be evaluated both from the perspective of teasing out tumor biology as well as identifying immune cells and their states.

The most common brain cancer types in children are embryonal cancers, including medulloblastomas. Previous genomic studies of human medulloblastomas have defined four consensus molecular subgroups (SHH, WNT, Group 3, and Group 4) based on discrete genomic and transcriptomic alterations within the tumors (Taylor et al., 2012). A recent investigation from a research group at St. Jude Children's Research Hospital analyzed single cell RNA-sequencing (scRNA-seq) from 25 pediatric medulloblastomas and revealed previously unknown differences between the subtypes of medulloblastomas, including cell types of origin from which the tumors arise within the brain (Hovestadt et al., 2019). However, the immune microenvironment in medulloblastomas remains understudied, especially in those children whose medulloblastomas recur following surgery and treatment, and thereby limits a long-term goal to implement immunotherapeutic strategies in the clinic.

Here, we present a study based on RNA expression of single cells obtained from pediatric medulloblastoma samples, demonstrating that these data can be utilized to characterize the subtype of the tumor and to evaluate immune gene expression and, by inference, the presence or absence of different immune cell types in pediatric CNS tumors. Our analysis includes the evaluation of data from single cell studies of RNA expression, to provide subtyping information

and to explore the presence of immune cell populations in primary and recurrent medulloblastomas. In particular, Table 1 provides information regarding the five medulloblastomas studied using scRNAseq (Table 1).

Table 1. Characteristics of the pediatric CNS cancer cohort used for scRNAseq.

Sample ID	Tumor timepoint	Diagnosis	Brain region	Sex	Age
CH79_P	primary	medulloblastoma group 4	fourth ventricle	female	12
CH79_R	recurrent	medulloblastoma group 4	frontal lobe		17
CH34	primary	medulloblastoma group 4	posterior fossa/fourth ventricle	male	8
CH161	primary	medulloblastoma group 4	posterior fossa	male	6
CH87	primary	medulloblastoma group 3	posterior fossa	male	6

MATERIALS AND METHODS

RNA extraction

RNA was extracted from frozen brain tumor tissue using mirVana isolation kit (Thermo Fisher Scientific) according to the manufacturer's protocol.

Immune gene expression assay and analysis

Multiplexed gene expression profiling was performed using the PanCancer Immune Profiling Panel (version 1.1) from NanoString Technologies, which assays expression of 770 immune-related and housekeeping genes. This platform provides gene-specific probes that carry a unique sequence and are hybridized to tumor-extracted DNA under optimal hybridization conditions overnight. Subsequent hybridization of this mixture with sequence-specific fluorescent-labeled oligonucleotides followed by imaging on the nCounter instrument, provides quantitative expression for each of the 770 genes in the assay. Briefly, tumor-extracted RNA (150 ng) was prepped according to manufacturer's protocol and analyzed on the nCounter instrument. Data were normalized using the housekeeping genes included in the panel. A category summary of the immune genes on the PanCancer panel is listed in Table 2. Differential expression was performed using the mean of each group to calculate a fold-change expression difference and by calculating an adjusted p-value (Benjamini false discovery rate) which corrects for multiple comparisons in a large, multiplexed gene expression dataset.

Table 2. Summary of PanCancer immune gene categories.

Category	# of Genes	Category	# of Genes
Chemokines	99	T-Cell Functions	72
Cytokines	56	Adhesion	25
Cell Functions	155	Complement	15
B-Cell Functions	18	Senescence	12
Antigen Processing	22	Interleukins	38

Regulation	154	Macrophage Functions	6
Cytotoxicity	10	TLR	11
NK Cell Functions	17	Microglial Functions	5
Transporter Functions	22	TNF Superfamily	30
Pathogen Defense	12	T-Cell Functions	72
Cell Cycle	13	Adhesion	25
Leukocyte Functions	6		

Nuclei isolation

Five samples of between 5-20 milligrams of surgically resected pediatric brain tumors underwent a nuclei isolation protocol (Table 2). Each sample was homogenized by douncing in nuclear isolation medium (NIM; sucrose: 0.25M, KCl: 25mM, $MgCl_2$: 5mM, TrisCl: 10mM), dithiothreitol, and 50x propidium iodide. Following douncing, 10% Triton-X was added, and the samples were stained with 2 μL of a nuclear stain (Hoescht), allowing for the visualization of nuclei during cell sorting. Nuclei were gated based on Hoescht signal and sorted into batches for a total of 15,000 nuclei that were subsequently added into a master mix containing reagents for single nuclei capture and library preparation. Reverse transcriptase (RT) enzyme was added to each sample. These samples were loaded on a 10 X Genomics Chromium Controller in order to generate single cell Gel Bead-In Emulsions (GEMs), utilizing the single cell 3'reagent kit v2 from 10x Genomics.

Library preparation and sequencing

Following preparation of the GEMs, single nuclei RNA sequencing libraries were prepared for next generation sequencing (NGS), following manufacturer instructions for the single cell 3'reagent kit v2 (10x Genomics). Briefly, the samples were amplified by 12 cycles of PCR during cDNA amplification based on a targeted recovery of >6,000 nuclei. After sample indexing, the final libraries underwent a number of PCR amplification cycles based on the cDNA

input (Table 3). Post-library construction quality control metrics were obtained using 1) Qubit quantitation to measure concentration of the final library and 2) Agilent TapeStation to evaluate the size of the final NGS library. Figure 1 shows a schematic of the final library product. Final libraries were pooled and sequenced as a five sample multiplex on an Illumina NovaSeq S1 flow cell. The run recipe is as follows: Read 1 – 28 cycles; i7 Index – 8 cycles; i5 Index – 0 cycles; Read 2 – 91 cycles. Table 3 provides the corresponding quality control and sequencing metrics.

Table 3. Single cell library prep and sequencing metrics.

Sample ID	Final library amplification cycles	Concentration (ng/uL)	Size (bp)	Cells sequenced	Mean reads per cell	Median genes per cell
CH79_P	14	14.5	562	9209	19540	1846
CH79_R	16	42.2	463	6513	51845	1460
CH34	14	11.7	525	12634	13297	1307
CH161	14	6	555	3715	54163	2117
CH87	14	8.44	444	9466	21398	1835

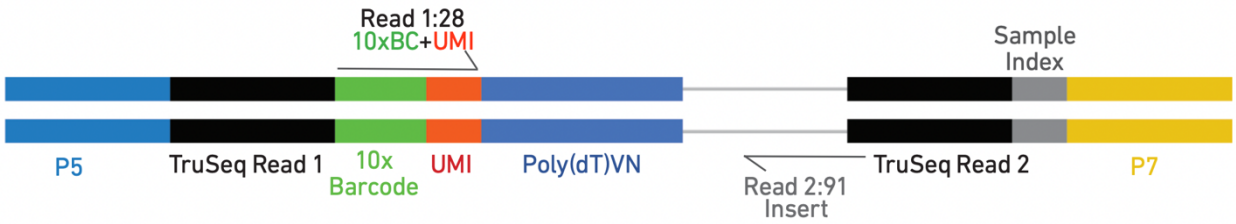


Figure 1. Schematic of final sequencing library. The Chromium Single Cell 3' Gene Expression library is flanked with paired-end Illumina primers P5 and P7 (blue and gold, respectively). Encoded in Read 1 is the 10x Genomics barcode (green) which is unique and specific to each captured cell and the Unique Molecular Identifier (UMI, red) sequence which is unique to each originating library fragment. TruSeq Read 1 and TruSeq Read 2 (black) are Illumina primer sites used to initiate forward and reverse sequencing reactions. Poly(dT)VN (violet) is a poly(T) site used to capture poly(A) tails found in most human mRNAs. Read 2

sequences the cDNA insert. Each sample is indexed (grey). (Figure obtained from 10x Genomics Chromium Controller User Manual).

Single nuclei sequencing analysis

Data pre-processing, including alignment, filtering, barcode-based sorting, and unique molecular identifier (UMI) counting, were performed using the 10x Genomics CellRanger software suite following the default parameters. Downstream analysis of gene expression was performed using Seurat v.3 for R (Satija et al., 2015). Briefly, the feature-barcode matrices were log-normalized and top variable genes across single cells were identified using the FindVariableGenes function. Data were scaled to 10,000 transcripts per cell. Dimensionality reduction was performed using principal component analysis (PCA) and then the distance matrix was organized into a K-nearest neighbor graph (KNN), partitioned into clusters using Louvain algorithm, and clusters were visualized on a Uniform Manifold Approximation and Projection (UMAP) plot. Top differentially expressed genes for each cluster were computed and cell types were annotated based on expression of known markers. The Human Protein Atlas was utilized to identify cell type specificity for the top differentially expressed marker genes for each cluster.

RESULTS

Medulloblastomas can be sub-grouped using gene expression data

Previous publications indicate that medulloblastomas separate into four distinct diagnostic subtypes. In particular, a subset of 22 genes can accurately characterize medulloblastoma subtypes (Northcott et al., 2012). Using a subset of pediatric medulloblastoma tumors from our cohort, including those we studied subsequently using scRNAseq, non-supervised hierarchical clustering analysis provided accurate subtype classification using gene expression data (Figure 2).

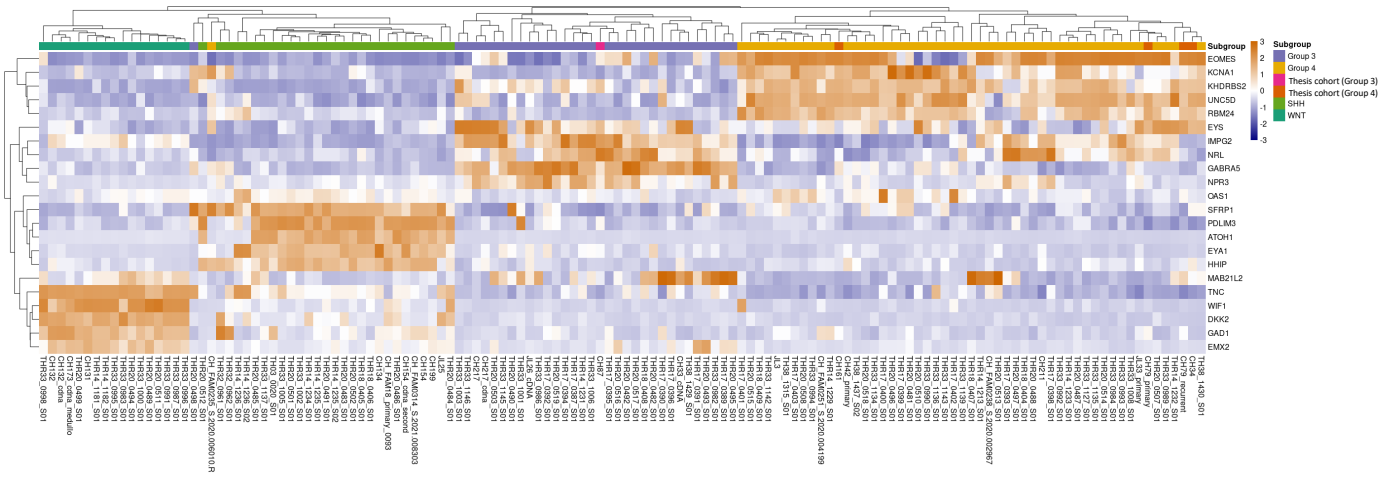


Figure 2. Heatmap demonstrating hierarchical clustering of medulloblastoma subgroup-specific gene expression signatures derived from bulk RNA sequencing. Relative gene expression levels are indicated according to the legend, where orange indicates higher relative expression compared to decreased relative expression in green. Each sample has a column and is defined above the heatmap as belonging either to group 3 (purple), group 4 (gold), SHH (light green), or WNT (dark green). Bulk RNAseq results from samples studied in the thesis cohort by scRNA sequencing are indicated as group 3 (pink), or thesis cohort group 4 (orange).

Sub-typing of medulloblastomas based on scRNA sequencing

When scRNAseq data from similar numbers of cells sequenced from a single group 3 medulloblastoma sample, CH87, were plotted in comparison to a single group 4 medulloblastoma sample, CH161, as shown in Figure 3, 10 distinct clusters were identified, many of which were sample specific. Clusters 0, 2, 4, and 7 were specific to the group 3 medulloblastoma sample (CH87); clusters 1, 3, 5, 6, and 10 were specific to the group 4 medulloblastoma sample (CH161); clusters 8 and 9 were shared between the two medulloblastoma subtypes. Genes previously identified to be specific to group 3 (Figure 4b) and to group 4 (Figure 4a) are plotted to show the specificity of clusters relative to medulloblastoma subtype.

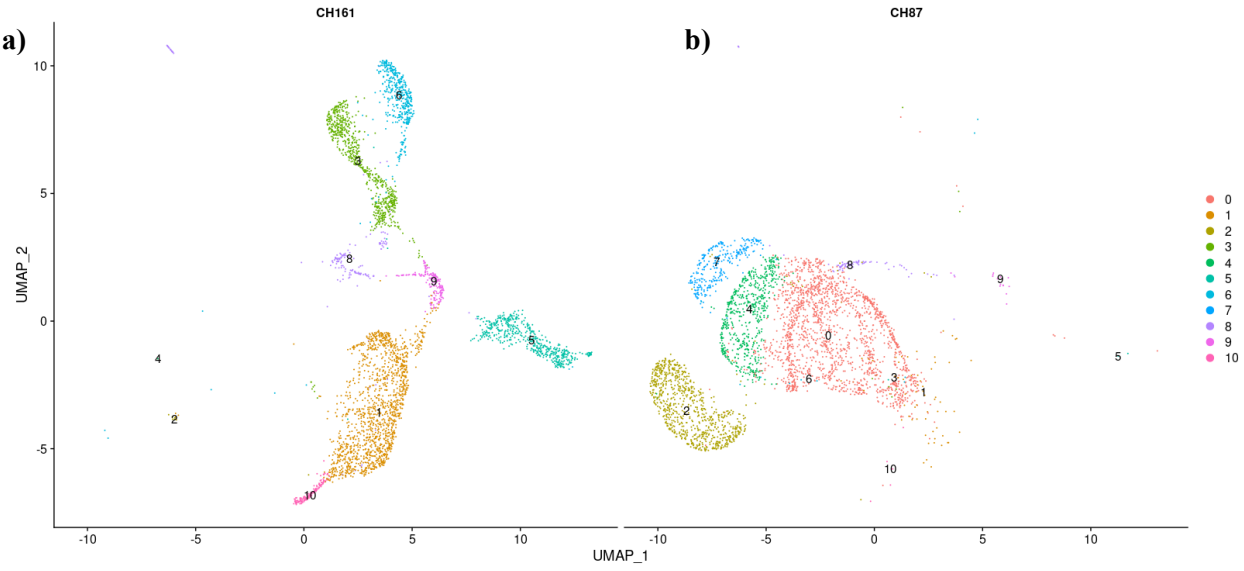
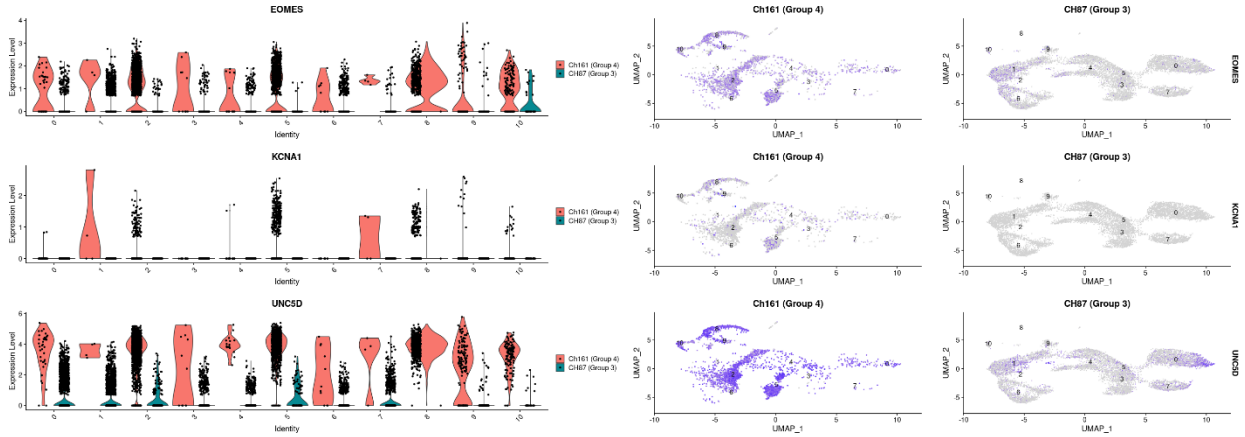


Figure 3. UMAP representation of gene expression clusters from a group 4 versus a group 3 medulloblastoma. a) Clustering of CH161, a group 4 medulloblastoma. **b)** Clustering of CH87 (randomly downsampled to 3,200 reads), a group 3 medulloblastoma.

a)



b)

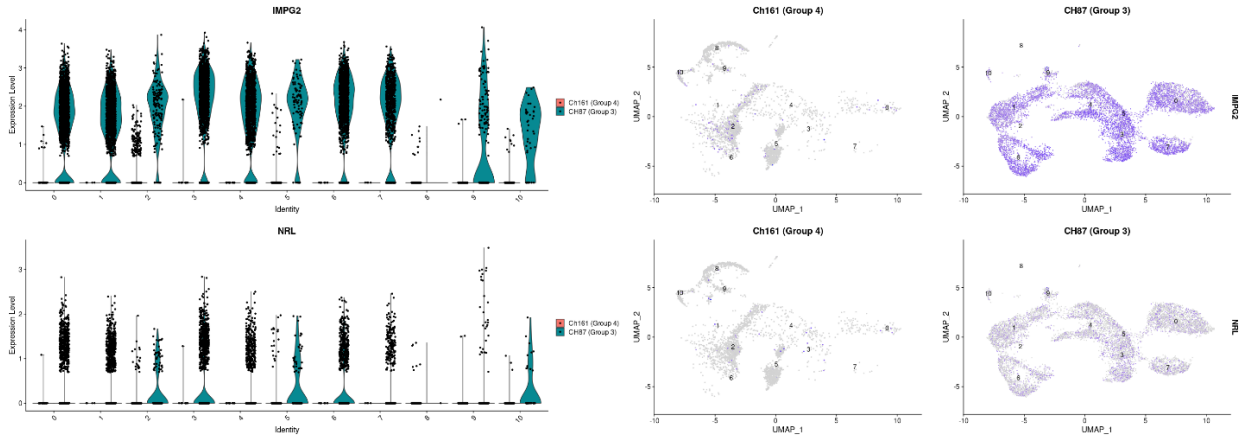


Figure 4. Expression levels of previously identified group 3- and group 4-specific genes in CH161 and CH87. Teal represents the expression of the analyzed marker genes in identified cell clusters of CH87 (group 3) and red represents the expression of the analyzed marker genes in identified cell clusters of CH161 (group 4) **a)** Expression of group 4 specific genes are plotted for CH161 and CH87. **B)** Expression of group 3 specific genes are plotted for CH161 and CH87.

Single nuclei sequencing clustering analysis and cell type identification

Five brain tissue tumor samples from pediatric medulloblastomas were studied by single nuclei capture and sequencing (Table 1). Our overall sequencing metrics revealed at least 3,715 individual nuclei were sequenced for each sample and the number of sequencing reads per cell/nuclei resulted in similar number of genes detected per cell (Table 3).

Gene expression data from all samples were analyzed together and clustered into 13 groups after processing in Seurat (Figure 5). Figure 5a displays all nuclei, colored by sample while Figure 5b displays the 13 clusters identified during our computational processing using UMAP for dimensionality reduction. By analyzing the top differentially expressed genes for each cluster compared to all other nuclei in the dataset, one can glean information about cell types and gene expression phenotypes. The top differentially expressed genes for each cluster are listed in Table 4.

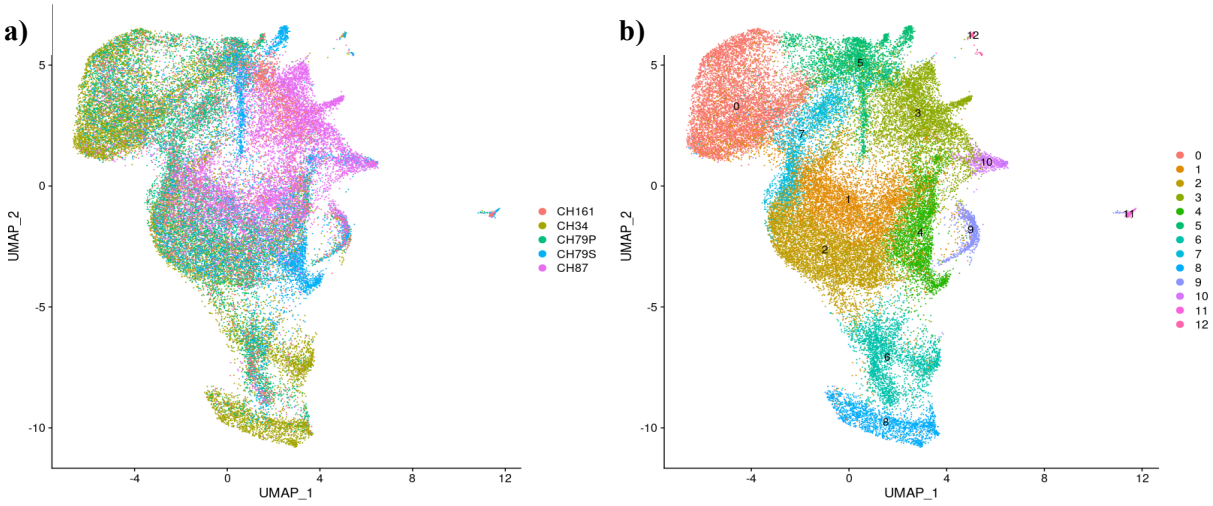


Figure 5. UMAP representation of the integrated single nuclei data set. a) Clustering separated by sample ID. **b)** Clustering separated by cell type.

Table 4. Top five differentially expressed marker genes in scRNAseq clusters.

Cluster	Gene	avg_logFC*	p_val_adj**	Cluster	Gene	avg_logFC*	p_val_adj**
0	KCNIP4	1.53938612	0	7	ESYT2	1.70449097	0
	DPYD	1.35681216	0		SEMA3E	1.51110781	0
	PDE1C	1.2574032	0		NTNG1	1.47652001	0
	MIR4300HG	1.18209031	0		AC092957.1	1.45478894	0
	LINC01632	1.1728838	0		NEGR1	1.44706014	0
1	NRG1	1.51222014	0	8	APOLD1	1.93534064	0
	SLC35F3	1.46712876	0		DIAPH3	1.84883511	0
	LRRTM4	1.31490884	0		TOP2A	1.56286466	0
	NLGN4Y	1.27169068	0		ASPM	1.48611701	0
	COL27A1	1.20953006	0		CENPF	1.42256865	0
2	CTTNBP2	1.27301983	0	9	PLCG2	2.0110145	4.13E-108
	GRIA2	1.17043542	0		CENPF	1.58727592	5.32E-27
	PTPRO	1.13826452	0		HSPH1	1.45021974	1.00E-117
	LINC00587	1.07644005	0		HSPA1A	1.30902929	1.26E-137
	ELP4	1.07131093	0		HIST1H2AC	1.27427086	2.24E-158
3	FAM19A4	1.93757924	0	10	NR2E3	2.21033028	0.00E+00
	AC009264.1	1.69480644	0		PTPN13	1.98060004	0
	TENM3	1.68101558	0		DISC1	1.93377412	0
	SYT1	1.67249429	0		CAMK1D	1.87211038	0
	CDH20	1.60140566	0		FKBP5	1.82298518	0
4	ADARB2	1.89407131	0	11	LRMDA	3.45642452	0
	CNTNAP2	1.71739521	0		SFMBT2	3.30195983	0
	GRIK1	1.4047383	0		SLC1A3	3.03753939	0
	CNTNAP3B	1.34340535	0		NEAT1	3.01980696	1.36E-185
	IQCJ-SCHIP1	1.31305695	2.18E-88		PLXDC2	2.96735626	2.38E-102
5	NTRK3	1.82287826	0	12	LRP1B	3.14392107	0
	GRM7	1.75361017	0		PCDH9	2.83901049	1.04E-162
	CTNNA2	1.7149246	0		KAZN	2.75793726	3.15E-153
	RIT2	1.63047899	0		ROBO2	2.59394379	1.64E-64
	PRKCE	1.43485037	0		NKAIN2	2.57743747	1.77E-168
6	BRIP1	1.52201886	0				
	DIAPH3	1.42045159	0				
	IL1RAPL1	1.38676748	0				
	HELLS	1.14709102	0				
	AP001347.1	1.13969168	0				

*avg_logFC log fold-change of the average expression between the cluster vs. all other cells

**p_val_adj adjusted p-value, based on Bonferroni correction using all genes in the dataset

Immune gene expression in a cohort of 121 pediatric brain tumors

A cohort of 121 pediatric CNS tumor RNA samples were subjected to immune gene expression analysis using the PanCancer Immune Panel. The cohort included: 16 medulloblastomas, 17 embryonal non-medulloblastoma tumors, 7 choroid plexus tumors, 14 diffuse astrocytic and oligodendroglial tumors, 25 other astrocytic tumors, 7 ependymal tumors, 3 germ cell tumors, 30 neuronal and mixed neuronal-glial tumors, and 2 tumors of the pineal region. For analysis, tumors were grouped into medulloblastomas (n = 16) vs. all other CNS tumors (n = 105). Differential expression analysis for each gene on the panel was performed to identify significantly differentially expressed genes within the medulloblastoma cohort. Our analysis revealed that several immune genes were significantly downregulated, or under expressed, in medulloblastomas compared to other CNS tumors (Figure 6). Some of the most downregulated immune genes in medulloblastomas were *ITGB4, IL1RAP, KLRC2,* and *IL1B*. Few immune genes were significantly overexpressed in medulloblastomas and include *SMPD3, ITGA2B, CXCR4,* and *CD24*.

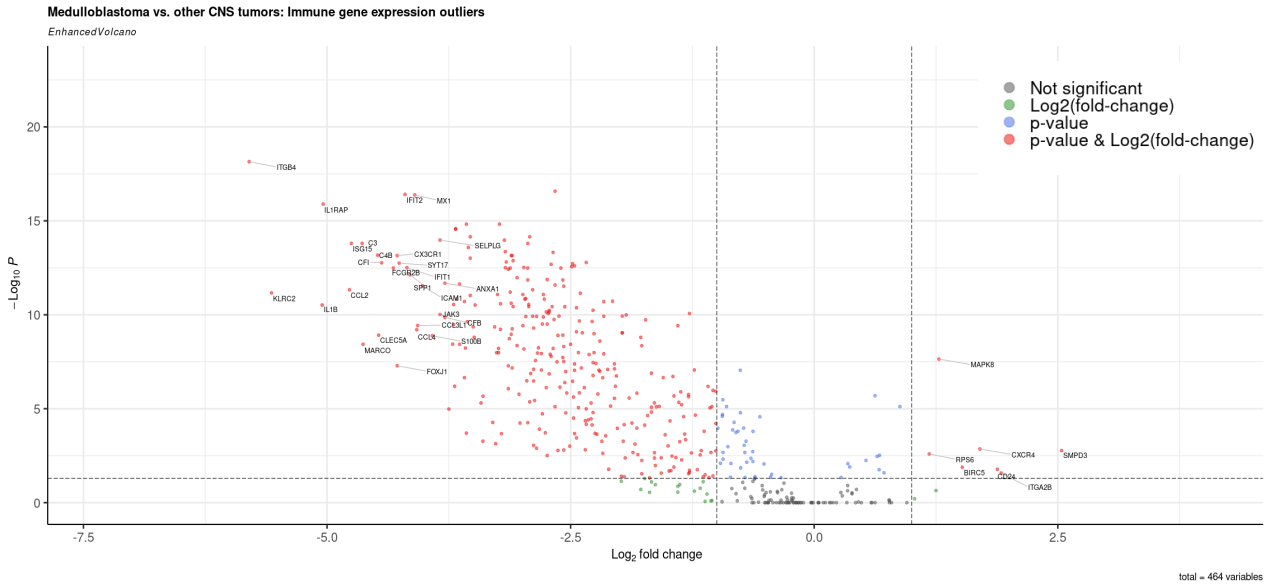


Figure 6. Volcano plot of immune gene expression outliers in medulloblastoma tumors versus other brain tumors. Gene expression values of the PanCancer immune genes in a cohort of medulloblastomas versus other brain tumors are plotted; each dot represents a single gene. The X axis displays the Log_2 fold change in gene expression, and the Y axis displays the P value from our differential expression analysis comparing medulloblastoma tumors to other brain tumors. P value cut off was set at 0.05 and fold change cut off at 2. Red marker dots indicate a significant Log_2 and p-value. Green marker dots indicate a significant fold-change value only, blue marker dots signify a significant p-value only, and grey marker dots are neither significant in fold-change nor p-value.

Analysis of primary versus recurrent medulloblastoma

Our cohort included a primary and recurrent medulloblastoma from the same patient ("CH79"), with a five-year duration between the recurrence and the initial diagnosis. Clustering of the scRNAseq data from each tumor and gene expression-based cell type analysis revealed that most clusters identified for each tumor were neurons or astrocytes (Figure 7a). One cluster, however, reflected gene expression in monocytes/macrophages and was present only in the recurrent tumor sample (Figure 7b). Certain differentially expressed genes in this cluster indicating monocytes/macrophages include *CD14, CD45*, and *MSR1*. This cluster also appeared to represent other immune cells, including naïve T-cells, due to expression of *PTPRC*, *FKBP5, MAML2, SNED1, PRKCA, and SLC40A1*.

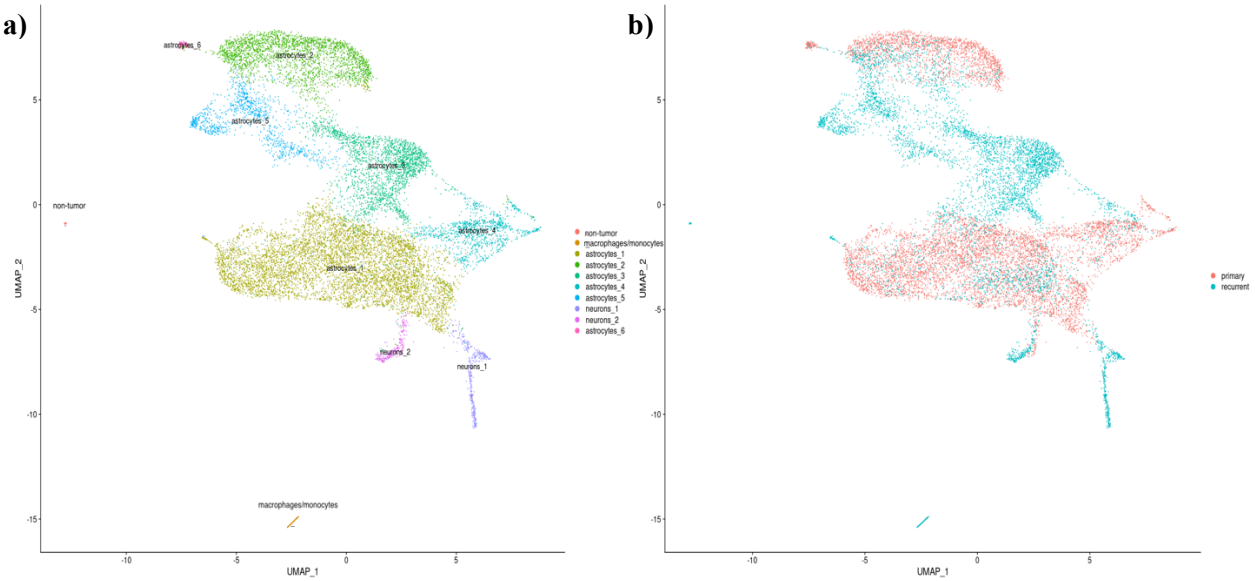


Figure 7. UMAP representation of a primary and recurrent medulloblastoma tumor. a) cell type identified clusters between both primary and recurrent medulloblastomas from CH79. **b)** primary tumor gene expression-based clustering (red) versus recurrent tumor clustering (teal) of the two longitudinal medulloblastoma samples indicates the macrophage/monocyte cluster is entirely unique to the recurrent sample.

Discussion

The characterization of medulloblastomas including different cell biology that defines subtypes of the disease, and its tumor microenvironment is vastly understudied. Through bulk RNA sequencing and scRNA sequencing, these aspects can be further defined in a robust manner. This work describes how characterization of the tumor microenvironment of the immune quiet medulloblastoma can be achieved through bulk RNA analysis and describes the way that scRNA sequencing can add precision to bulk RNA sequencing findings.

First, proper subtyping of medulloblastomas based on their RNA expression is possible. Bulk RNA sequencing can be utilized in the characterization of medulloblastomas, as exhibited in Figure 2, through gene expression-based clustering. The use of scRNA sequencing was also utilized in the identification of clusters and cellular expression profiles of group 3 versus group 4 medulloblastomas. This technique can identify clusters of nuclei that are unique to or shared by the two groups. Figure 3 shows that CH87 (group 3) and CH161 (group 4) differ in how their gene expression data from nuclei are clustered, which infers differences in biology of these subtypes that contribute to their prognosis. This method of subtyping can be supported by determining whether genes that were previously identified to be specific to either group 3 or group 4 medulloblastomas cluster with their respective sample. These sub-group marker genes were highlighted in each sample in Figures 6a and 6b, demonstrating known medulloblastoma genes specific to group 3 are most highly expressed in CH87 and genes specific to group 4 are most highly expressed in CH161. These data indicate that subtyping of medulloblastomas through scRNA sequencing is possible and follows pathologically predicted subtyping. With additional samples studied, we may be able to identify the specific biology of different cell type

clusters that impart known prognosis on these different subtypes, for example those cells having stem-like gene expression that may persist in recurrent disease.

It is also important to note that Figure 2 begins to identify the heterogeneity of medulloblastomas. Two 'outliers' did not accurately cluster with their respective pathologically defined subtypes, showing that not every cancer fits into a perfectly defined set of rules and that cancers have a way of inventing their own expression and are biologically unique. Figure 5 aids in identifying unique cell populations of medulloblastomas and their expression profiles. The top differentially expressed genes for each cluster can be derived from these data and applied in many settings, such as understanding the expression profile of a certain subset of cells in a tumor that may not categorize well into a specific subtype of medulloblastoma.

Indeed, some of these unique cell clusters may represent immune cell components of the tumor mass. To identify these by gene expression we studied a large number of CNS cancers including medulloblastomas to identify unique over- and under-expressed genes in medulloblastomas. Here, gene expression analysis using the PanCancer Immune Panel yielded a volcano plot (Figure 6) that exhibited decreased expression of immune genes such as *KLRC2,* a gene expressed in natural killer cells, and *IL1B,* a gene that encodes a cytokine involved in inflammation. The volcano plot indicates that medulloblastomas primarily exhibit a decreased expression of immune genes, though a smaller population of genes were overexpressed and include *CD24*. Although *CD24* modulates B-cell activation responses, there is also evidence that it is highly expressed in developing pediatric brain (*Reactome*). From these data, one can begin characterizing the immune cell content of medulloblastomas as low compared to other CNS tumors. Through the scRNA sequencing of longitudinal samples, namely primary versus recurrent medulloblastomas, this characterization can be further refined. Figure 7b indicates

differing cell expression profiles between a primary and recurrent medulloblastoma from the same patient. Notably, the recurrent medulloblastoma of CH79 differentially expresses *CD14, CD45*, and *MSR1* which are indicators of monocytes/macrophages. This may indicate that immune infiltration of monocytes/macrophages is higher in recurrent medulloblastomas, which can only be substantiated by single cell analyses of additional primary/recurrent pairs due to the low level of infiltration evidenced.

Through characterization of the tumor microenvironment and expression profile of different cell types present in medulloblastomas, as well as the comparison of the expression profile between different subtypes and primary versus recurrent tumors, better outcomes for medulloblastoma patients could be possible in the future. Looking forward, future clinical research ought to include immunotherapies that act to recruit immune cells into the microenvironment, such as oncolytic virotherapy given these findings of relatively low immune infiltration in medulloblastomas.

References

Curtin, S. C., Minino, A. M., & Anderson, R. N. (2016). Declines in Cancer Death Rates Among Children and Adolescents in the United States, 1999-2014. *NCHS Data Brief*, *257*, 1–8.

Dagogo-Jack, I., & Shaw, A. T. (2018). Tumour heterogeneity and resistance to cancer therapies. *Nature Reviews Clinical Oncology*, *15*(2), 81–94. https://doi.org/10.1038/nrclinonc.2017.166

Gröbner, S. N., Worst, B. C., Weischenfeldt, J., Buchhalter, I., Kleinheinz, K., Rudneva, V. A., Johann, P. D., Balasubramanian, G. P., Segura-Wang, M., Brabetz, S., Bender, S., Hutter, B., Sturm, D., Pfaff, E., Hübschmann, D., Zipprich, G., Heinold, M., Eils, J., Lawerenz, C., … Pfister, S. M. (2018). The landscape of genomic alterations across childhood cancers. *Nature*, *555*(7696), 321–327. https://doi.org/10.1038/nature25480

Hovestadt, V., Smith, K. S., Bihannic, L., Filbin, M. G., Shaw, M. L., Baumgartner, A., DeWitt, J. C., Groves, A., Mayr, L., Weisman, H. R., Richman, A. R., Shore, M. E., Goumnerova, L., Rosencrance, C., Carter, R. A., Phoenix, T. N., Hadley, J. L., Tong, Y., Houston, J., … Northcott, P. A. (2019). Resolving medulloblastoma cellular architecture by single-cell genomics. *Nature*, *572*(7767), 74–79. https://doi.org/10.1038/s41586-019-1434-6

Ma, X., Liu, Y., Liu, Y., Alexandrov, L. B., Edmonson, M. N., Gawad, C., Zhou, X., Li, Y., Rusch, M. C., Easton, J., Huether, R., Gonzalez-Pena, V., Wilkinson, M. R., Hermida, L. C., Davis, S., Sioson, E., Pounds, S., Cao, X., Ries, R. E., … Zhang, J. (2018). Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature*, *555*(7696), 371–376. https://doi.org/10.1038/nature25795

Northcott, P. A., Shih, D. J. H., Remke, M., Cho, Y.-J., Kool, M., Hawkins, C., Eberhart, C. G., Dubuc, A., Guettouche, T., Cardentey, Y., Bouffet, E., Pomeroy, S. L., Marra, M.,

Malkin, D., Rutka, J. T., Korshunov, A., Pfister, S., & Taylor, M. D. (2012). Rapid, reliable, and reproducible molecular sub-grouping of clinical medulloblastoma samples. *Acta Neuropathologica*, *123*(4), 615–626. https://doi.org/10.1007/s00401-011-0899-7

Reactome | CD24(29-534) [plasma membrane]. (n.d.). Retrieved April 16, 2021, from https://reactome.org/content/detail/R-I-430062

Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., & Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nature Biotechnology*, *33*(5), 495–502. https://doi.org/10.1038/nbt.3192

Sharma, P., & Allison, J. P. (2015). The future of immune checkpoint therapy. *Science*, *348*(6230), 56–61. https://doi.org/10.1126/science.aaa8172

Taylor, M. D., Northcott, P. A., Korshunov, A., Remke, M., Cho, Y.-J., Clifford, S. C., Eberhart, C. G., Parsons, D. W., Rutkowski, S., Gajjar, A., Ellison, D. W., Lichter, P., Gilbertson, R. J., Pomeroy, S. L., Kool, M., & Pfister, S. M. (2012). Molecular subgroups of medulloblastoma: The current consensus. *Acta Neuropathologica*, *123*(4), 465–472. https://doi.org/10.1007/s00401-011-0922-z

Venteicher, A. S., Tirosh, I., Hebert, C., Yizhak, K., Neftel, C., Filbin, M. G., Hovestadt, V., Escalante, L. E., Shaw, M. L., Rodman, C., Gillespie, S. M., Dionne, D., Luo, C. C., Ravichandran, H., Mylvaganam, R., Mount, C., Onozato, M. L., Nahed, B. V., Wakimoto, H., … Suvà, M. L. (2017). Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science*, *355*(6332). https://doi.org/10.1126/science.aai8478

Acknowledgements

I would like to thank my mentor and research advisor, Dr. Elaine Mardis, as well as my mentor, Dr. Katherine Miller, for all of their advice, support, and knowledge over the last two years that I have had the honor of working with them. I would also like to thank all members of the Mardis lab as well as those that work for the Institute for Genomic Medicine at Nationwide Children's Hospital for their tireless support. I would like to thank my advisor, Dr. Gregory Booton, for his guidance through my journey at The Ohio State University.

I would also like to thank The Ohio State University's College of Arts and Sciences for supporting this research by granting me the Undergraduate Research Scholarship in the spring of 2020.